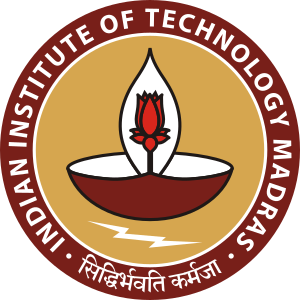
**Data Analytics & Business Insights Report**



**Team 12 LumiGraph**

**Prepared by –**

Shivendra Patel (21f2001310)

Banoth Pavan (21f1002144)

Neeraj Yadav (21f3002261)

Thoomati Aditya Vikram (21f1000610)

Contents

[1.Overview - Purpose & Scope 1](#_Toc194749437)

[2. DataFrames Overview 1](#_Toc194749438)

[3. Data Cleaning Steps 3](#_Toc194749439)

[4. Summary of Changes 5](#_Toc194749440)

[5. Strategic Insights & Recommendations 6](#_Toc194749441)

[6. Dashboard Analysis 9](#_Toc194749442)

[7. Conclusion 9](#_Toc194749443)

# 1.Overview - Purpose & Scope

This report documents **end-to-end data preprocessing** and **analytical insights** derived from:

* **Sales Data**: Amazon Sales, International Sales, and Sales Reports.
* **Operational Data**: Warehouse records (Pricing, KPIs, Contracts).
* **Financial Data**: P&L Statements (March 2021 & May 2022)

Link for Dashboard - [Dashboard Link](https://public.tableau.com/shared/TZMPNQQ36?:display_count=n&:origin=viz_share_link)

Link for Presentation - [Presentation Link](https://docs.google.com/presentation/d/1SqH48KpZKEYt7OBLHmGGVwCH5OQGDWcwK5pel1k92hI/edit?usp=sharing)

Link for Kaggle Notebook Analysis - [Python Analysis](https://www.kaggle.com/code/shvndrptl/dv-project-eda-lumigraph)

# 2. DataFrames Overview

**2.1 Amazon Sales Report**

* **Key Columns**:
  + Order\_ID
  + Order\_Date
  + Order\_Status
  + Fulfillment\_Type
  + Sales\_Channel
  + Product\_Style
  + Product\_SKU
  + Quantity
  + Sale\_Amount
  + Shipping\_City
  + Shipping\_State

**2.2 International Sales Report**

* **Key Columns**:
  + Sale\_Date
  + Sale\_Month
  + Customer\_Name
  + Product\_Style
  + Product\_SKU
  + Quantity\_Purchased
  + Price\_per\_Unit
  + Gross\_Amount

**2.3 Warehouse Data**

* **Key Sections**:
  + Pricing Data
  + Operational Processes
  + Contractual Terms (Clauses and Requirements)
  + Key Performance Indicators (KPIs)

**2.4 Sales Report (sales\_df)**

* **Key Columns**:
  + Product\_SKU
  + Sale\_Amount
  + Quantity
  + Other relevant fields

**2.5 PandL Datasets**

* **March 2021**: No missing values or inconsistencies identified.
* **May 2022**: No missing values or inconsistencies identified.

# 3. Data Cleaning Steps

**3.1 Amazon Sales Report**

**Step A: Handle Cancelled Orders**

* Set Sale\_Amount to zero for orders with a Courier\_Status of "Cancelled".
* Updated the Currency field to "INR" for cancelled orders.

**Step B: Handle Empty Courier\_Status**

* Filled empty values in Courier\_Status with "Adjustment" to maintain consistency. We did not fill it with cancellations as we are not sure about it so we filled it with adjustment which we can change latter to cancelled.

**Step C: Fill Missing Currency Values**

* Filled missing values in the Currency column with "INR".

**Step D: Verify and Handle Remaining Missing Sale\_Amount Values**

* Checked for any remaining missing values in Sale\_Amount.
* Filled any remaining missing Sale\_Amount values with zero, assuming they are cancelled or 0 transactions.

**Step E: Verify Changes and Unique Values**

* Reviewed unique values in the Courier\_Status column to ensure all statuses were appropriately categorized.
* Checked for any remaining missing values after cleaning.

**Step F: Basic Statistics After Cleaning**

* Generated basic statistics for the Sale\_Amount column to assess data distribution.

**Step G: Count of Orders by Courier\_Status**

* Counted the number of orders categorized by Courier\_Status to understand order fulfillment performance.

**3.2 International Sales Report**

**Step A: Handle Missing Values**

* Identified and filled missing values in key columns such as Gross\_Amount.

**Step B: Correct Data Formats**

* Converted Sale\_Date from string format to a proper date format (YYYY-MM-DD).
  + Converted Sale\_Date to datetime format to enable date-related operations.

**Step C: Fill Missing Sale\_Month Values:**

* Checked for missing values in the Sale\_Month column (36,391 missing).
  + Filled missing values in Sale\_Month based on corresponding entries in Sale\_Date.
  + Used .dt.month\_name() after ensuring that Sale\_Date was converted to datetime format.

**Step D: Standardize Product Information**

* Standardized product SKUs and sizes to ensure consistency across entries (e.g., ensuring consistent naming conventions).

**3.3 Warehouse Data**

**Pricing Data Cleaning:**

1. Extracted relevant pricing data from specific rows.
2. Dropped redundant index columns.
3. Renamed columns for clarity.
4. Removed header rows that contained unnecessary information.
5. Cleaned price data by removing currency symbols and converting them to numeric types.

**Operational Processes Cleaning:**

1. Extracted relevant rows for operational processes.
2. Renamed columns based on their actual names.
3. Forward-filled missing process stage information.
4. Dropped rows with completely missing sub-process information.

**Contractual Terms Cleaning:**

1. Extracted clauses and requirements from specified rows.
2. Forward-filled section headers to maintain context.
3. Cleaned clause text to remove non-alphabetic characters.

**KPI Table Cleaning:**

1. Extracted relevant KPI data from specified rows.
2. Dropped rows with missing KPI information.

**Miscellaneous Section Cleaning:**

1. Extracted miscellaneous clauses and requirements.
2. Forward-filled section headers.
3. Cleaned clause text similarly to contractual terms.

**3.4 Sales Report**

**Step A: Drop Redundant Columns**

* Removed the redundant index column from the dataset.

**Step B: Remove Rows with Missing Product\_SKU Values**

* Dropped rows where Product\_SKU, a critical identifier, was missing to ensure data integrity.

**3.5 PandL Datasets**

**March 2021 & May 2022:**

Both datasets were reviewed thoroughly, and no missing values or inconsistencies were identified across any columns or rows. As such, no cleaning steps were required for these datasets.

# 4. Summary of Changes

**Amazon Sales Report:**

1. Handled cancelled orders by setting their sale amounts to zero and updating currency fields.
2. Addressed empty values in critical columns like Courier\_Status.
3. Filled missing currency values with "INR".
4. Verified changes and ensured no critical missing values remained.

**International Sales Report:**

1. Filled missing values in key columns like Gross\_Amount.
2. Converted date formats for consistency.
3. Filled missing Sale\_Month values based on corresponding entries in Sale\_Date.
4. Standardized product SKUs and sizes for uniformity across entries.

**Warehouse Data:**

1. Cleaned pricing data by removing unnecessary headers and formatting prices correctly.
2. Structured operational processes with clear naming conventions.
3. Organized contractual terms and KPIs for clarity and usability.

**Sales Report:**

1. Removed redundant index column.
2. Dropped rows with missing critical identifiers (Product\_SKU).

**PandL Datasets:**

No cleaning was required as both datasets were complete and consistent.

**Conclusion**

The data cleaning processes ensured that all datasets are free of missing values, have consistent formats, and are standardized for analysis purposes across platforms, sales channels, and operational processes.

# 5. Strategic Insights & Recommendations

**5.1. Logistics & Fulfillment Strategy**

**Insight:**

* Shiprocket offers **predictable costs** (lower outbound/returns); INCREFF suits short-term storage needs.

**Recommendations:**  
 **Adopt Hybrid Logistics:**

* Use **Shiprocket for high-return categories** (e.g., Western Dress) to cut costs.
* Partner with **INCREFF for flash sales/short-term storage** (e.g., seasonal Dupattas in June).  
   **Negotiate Contracts:**
* Leverage Shiprocket’s stability for **Kurta/Kurta Sets** (high volume).

**5.2. Inventory & Sales Optimization**

**Insights:**

* **Fast-Movers:** Kurta (volume) & Kurta Set (revenue) dominate.
* **High-Margin:** Gown/Tops yield higher profit despite lower volume.
* **Stock-Sales Link:** Higher stock of Kurta/Kurta Sets → Higher sales.

**Recommendations:**  
 **Dynamic Stock Allocation:**

* **Boost inventory** of Kurta/Kurta Sets (year-round) + **Gowns/Tops** (high-margin).
* **Reduce Bottoms post-May** (seasonal decline).  
   **Promote Bundles:**
* Pair **Kurta Sets with Dupattas** in June (festive synergy).

**5.3. Shipping & Cancellation Analysis**

**Insights:**

* **Expedited > Standard Orders:** Customers prefer speed (Amazon’s fast delivery).
* **Cancellations:** Higher in merchant-fulfilled (slow delivery) & non-promo orders.

**Recommendations:**  
 **Prioritize Amazon Fulfillment:**

* Use Amazon for **low-value, fast-moving items** (Kurta) to reduce cancellations.

**Run Promotions Strategically:**

* Offer **discounts on merchant-fulfilled orders** to offset cancellation risks.

**5.4. Seasonal & Marketing Strategy**

**Insights:**

* **Peak Season:** Sets peak in April; Western Dress/Tops in months 4–5 (summer/weddings).
* **Color Trends:** Blue, Green, Pink drive sales.

**Recommendations:**  
 **Time-Blocked Campaigns:**

* **March:** Launch pre-peak campaigns for **Sets**.
* **April–May:** Push **Western Dress/Tops** (summer collections).  
  **Color-Centric Marketing:**
* Highlight **Mustard/Peach** in summer; **Blue/Green** year-round.

**5.5. Profitability Focus**

**Insight:**

* **Gowns/Tops:** Lower volume but higher margins.

**Recommendation:**  
 **Premium Positioning:**

* Market Gowns/Tops as **luxury/occasion wear** with higher price points.

**5.6. Cancellation Mitigation**

**Action Plan:**

* **Promotions:** Bundle with high-demand items (e.g., “Free Dupatta with Kurta Set”).
* **Delivery Speed:** Partner with **express logistics** for merchant orders.

# 6. Dashboard Analysis

**Key Features**

1. **Dynamic Analysis**

* Avg. Margin: [avg(Profit\_Margin\_Percentage)]%
* Filters: Product Category | Time Period

1. **Interactive Visualizations**

* Dual-Axis Charts: Compare sales vs. cancellation rates
* Geographic Heatmaps: State-wise revenue distribution
* Sales Trends: Toggle between daily/weekly/monthly views
* Pie And Bubble Charts for Order and Courier Status

**Dashboard Guide:**

1. **Select** a product category from the dropdown, then **toggle** between weekly/monthly views using the time selector.
2. **Hover over** bars to see details related to any chart.

# 7. Conclusion

Our analysis reveals three key opportunities:

1. Cost Reduction: Through optimized logistics
2. Revenue Growth: Via seasonal bundling
3. Customer Retention: By reducing cancellations

**Final Takeaway:**  
Align inventory, logistics, and marketing with **seasonality** and **customer preferences** (speed, promotions) to maximize revenue and minimize cancellations.